

A longitudinal study of mathematical word problem solving in children using a computer-based metacognitive strategy

Hidetsugu Tajika¹⁾, Narao Nakatsu²⁾, Ewald Neumann³⁾, Hisae Kato⁴⁾,
Tomoko Fujitani⁵⁾, Chie Hotta⁶⁾, and Hironari Nozaki²⁾

The purpose of the study was to examine the efficacy of self-explanation for helping elementary school children solve mathematical word problems through computer-based support over two years. Twenty participants received training and testing from the fourth grade to the sixth grade. The students solved worked-out examples for thirty minutes once a week in six two-week training sessions. They completed a word problem test at the end of each session. The results showed that all of the students gradually solved more word problems correctly after training than before. We classified students in the training condition into three groups according to the patterns of their test scores from an initial test in February and a later one in July. Students in the upper group, who had consistently higher scores, appropriately chose self-explanation sentences of their solution processes using inferences. Some of the students in the middle group, who were gradually increasing their scores, also chose self-explanation sentences using inferences. Most students in the lower group, who had consistently low scores, did not choose self-explanation sentences using inference at first, but correctly selected them after receiving the feedback. Self-explanation as a metacognitive strategy is discussed

Key words : Self-explanation, Mathematical word problem solving, Computer-based support, Children, Longitudinal study

1 Introduction

The purpose of this study was to test the long-term efficacy of self-explanation as a metacognitive strategy for elementary school students learning to solve mathematical word problems through computer-based cognitive support over two years.

Many researchers have focused on metacognitive strategies that facilitate knowledge construction as a way to get students to engage in problem solving with greater understanding (Flavell, 1979; Palincsar & Brown, 1984; Schoenfeld, 1985, 1987, 1992).

For example, Schoenfeld (1992) showed that a metacognitive strategy such as a heuristic technique fostered university students' mathematical problem solving. University students solved nonroutine problems, after they had been trained to use a

variety of metacognitive strategies of monitoring and control, for example, analyzing a problem, searching it, planning, and reconfirming the result. As a result, such metacognitive strategies improved students' problem solving abilities. Palincsar and Brown also showed that the training of metacognitive strategies such as asking questions, summarizing, clarifying, and predicting improved students' reading comprehension. It is well known that there are a variety of metacognitive strategies, for example, self-questioning, asking questions, answering questions, summarizing, heuristics of problems, and so on.

Recent research has shown that self-explanation is an effective metacognitive strategy across a wide range of task domains (Chi & Wylie, 2014; Chiu & Chi, 2014). Self-explanations are statements a problem solver makes to

1) Kobe Shinwa Women's University 2) Aichi University of Education 3) University of Canterbury, New Zealand
4) Hyogo University of Education 5) Mukogawa Women's University 6) Kansai University of Welfare Sciences

themselves to convey information with the goal of making some contents clear and understandable. Self-explanations involve generating comments that contain domain-relevant information and sometimes contain contents beyond the information given. When students self-explain a worked-out example, they make sense of new information by using prior knowledge and inferences to explain it to themselves (Chi, 2000). For example, Chi, Bassok, Lewis, Reimann, and Glaser (1989) analyzed results from university students who had self-explained worked-out examples of physics problems. The result showed that better problem solvers made more inferential statements.

A number of studies have shown that students generally learn better when they explain tasks such as expository texts and physics problems to themselves and when they explain their own problem-solving steps (Bielaczyc, Pirolli, & Brown, 1995; Chi et al., 1989; Neuman & Schwarz, 1998, 2000; Tajika, Nakatsu, Nozaki, Neumann, and Maruno, 2007). For example, Bielaczyc et al. trained university students to self-explain while learning computer science. As a result, students in the self-explanation training group explained more functions in computer science and performed better. Tajika et al. used elementary school students to examine the effect of self-explanation. Sixth grade students self-explained each solution step to the worked-out examples of mathematical word problems using paper and pencil. Each worked-out example had six to nine solution steps. Students who self-explained each solution step performed better than those who only solved the word problems. The findings produced clear evidence supporting the effectiveness of self-explanation.

Recent research has examined the use of computer-based support to promote metacognitive strategies (e.g., Azevedo & Allevi, 2013). Our framework for self-explanation was also designed to solve each step for worked-out examples by providing computer-based support for self-explanation.

Within the realm of computer-based self-

explanation support systems, for example, Alevi and Koedinger (2002) used geometry problems to compare self-explanations emphasizing computer-based instructional environments to instructional methods that did not emphasize self-explanations. The results showed that tenth grade students who explained the solution steps of a geometry problem during problem-solving practice in computer-based instructional environments learned with greater understanding compared with students who did not explain their solution steps.

As stated above, research on computer-based support for self-explanation has mainly used high school students and university students as participants, or people taking training classes. We developed a computer-based support program that helps fourth grade elementary school children self-explain solution steps for worked-out examples of mathematical word problems over two years. Prior to the present study, Tajika, Nakatsu, Neumann, Nozaki, Kato, Fujitani, and Hotta (2012) reported the results for mathematical word problem solving using computer-based self-explanation support with fifth graders over one year.

In Tajika et al. (2012), seventy-one fifth graders participated in the study in the training condition. Students in the training condition were given computer-based self-explanation support training. They had four three-week training sessions during one year. After each session of the computer-based self-explanation support training, students took a mathematical word problem test. They also took a transfer test at the final session. Students in the control condition had the same mathematical word problem tests as those in the training condition two times and the same transfer test at the same time. The results showed that students in the training condition outperformed those in the control condition on the fourth mathematical word problem test and the transfer test.

The present study attempts to extend the results of Tajika et al. (2012) in a longitudinal study using fourth graders.

2 Method

2-1 Participants

Twenty fourth-grade students in an elementary school (13 girls and 7 boys with a mean age of 10 years 6 months when the study began) participated in the study from February 2012 to December 2013. In Japan an elementary school year has three terms; the first term (April to August), second term (September to December), and third term (January to March), so students participated in the study from the third term of fourth grade until the second term of the sixth grade. We also used the control condition data from Tajika et al. (2012) to compare with the results of the training condition using the twenty fourth-grade students mentioned above. The number of students in the control condition was 62 fifth graders from a different elementary school. Students in the control condition took a pretest in April. They just became fifth graders. The pretest of the control condition was carried out two months after the students in the training condition had finished the pretest.

2-2 Materials

We used 30 worked-out examples as the computer-based training word problems. Each example had five to nine solution steps and the answer. Half of the training word problems were easy and the other half were difficult. An easy word problem had one or two mathematical expressions, and a difficult one had more than three mathematical expressions. These worked-out examples involved target items (correct answers) and distractor options (errors). Students selected one of the answers (i.e., a target item) to indicate an explanation. For example, one of the worked-out examples, which had five solution steps and the answer, was as follows: “The science club had a capacity of 30 students at the elementary school. The ratio of students who hope to become members of the science club is $.6$. What is the number of students who hope to become members of the science club at the school?” Step 1 of the

worked-out example was as follows: “The science club has a capacity of 30 students at the elementary school. Please select the same meaning as the sentence. Choice items: 1. The science club limits the number of students to thirty. 2. The science club has 30 students.” Each solution step had two to four choice items.

The word problem tests used during the two-year longitudinal study consisted of one word problem pretest, six kinds of mathematical word problem tests, and one transfer test. The pretest consisted of two kinds of elimination problems and two kinds of decimal problems, which they had learned in the fourth grade. The mathematical word problem tests consisted of eight word problems. Each type of word problem test consisted of an easy word problem test and a difficult word problem test. Each of the eight mathematical word problem tests consisted of four easy word problems and four difficult word problems, whose problem types were similar to those of the word problem pretest. For example, one of the elimination problem for the fifth graders was as follows, “When you get to the roller coaster in an amusement park, one adult costs twice as much money as one child costs. Two adults and three children cost 2100 yen. How much does each of one adult and one child cost?” One of the decimal problem for the fifth graders was as follows, “You have 1.6 liters of honey, which weighs 2.4 kilograms. How much does 3 liters of honey weigh?” Students received a different word problem test for students in the sixth grade included ratio word problems in addition to the above two types of mathematical word problems.

The transfer test consisted of an 18-item word problem test, adapted from a multiple-choice test used by Mayer, Tajika, and Stanley (1991). It had three kinds of questions. One kind of question was to make a number sentence from a sentence such as, “Taro has 5 more apples than H anako”. Another kind of question was to write down the numbers to be needed to solve such a problem as, “Masao had 500 yen to buy lunch. He

bought a sandwich for 290 yen, an apple for 70 yen, and a pint of milk for 110 yen. How much money did he spend?" The other kind of question was to write down the operations to be carried out to solve such a problem as, "If it costs 100 yen per hour to rent a pair of roller skates, what is the cost of using the pair of skates from 1:00 p.m. to 3:00 p.m.?" Students were asked to generate an answer to each question instead of being given a multiple-choice test.

In addition to the word problem tests, students also self-explained each solution step to one worked-out example using a paper and pencil test after the computer-based training session in each term. Six of the worked-out examples used in the paper and pencil tasks were difficult word problems. These worked-out examples had seven to nine solution steps and blank spaces for self-explanations. Before students wrote down their explanations in the blank spaces on the example sheet, they were asked about whether they understood a problem solution at each step. When they understood it, they answered "yes" and then wrote down their explanations. When students did not understand the problem solution at each step, they circled the "no" answer in pencil. Then, they were encouraged to write down their explanations about why they did not understand it.

2-3 Procedure

The study had four sessions in the training condition. In the first session, all of the fourth grade students received a word problem pretest in February. The pretest consisted of four word problems and took 20 minutes.

In the second session, the students worked with computer-based self-explanation support training. The computer-based self-explanation training was conducted in the context of a computer class in the computer room of the elementary school. Each student was seated at an individual desktop computer and was instructed to work independently from their peers.

Our computer-based support program for self-explanation was designed to be accessible by a

Web server. The computer-based self-explanation program was developed from a Java applet and it runs on any computer with the JRE 1.6 or later version (JRE is the Java Runtime Environment). The students selected one of the worked-out examples displayed on a computer and solved it. Each correct solution which they selected was shown on a computer display. When students finished solving one of the worked-out examples, they described correct solutions on a computer display. They had six two-week training sessions up until December 2013 (i.e., February 2012, July 2012, December 2012, February 2013, July 2013, and December 2013). The students were asked to solve two or three worked-out examples for twenty minutes. After the training, each student then described explanations using his or her own words on a notebook, which was given by the experimenter before the training. The students wrote the correct responses in their notebooks, which were shown on the computer display. It took five minutes.

In the third session, the students received two kinds of tests. The first test was called a paper and pencil test of worked-out examples. The students self-explained each solution step to one worked-out example using a paper and pencil test after the computer-based training session. They took the paper and pencil test for ten minutes. The second test was conducted one week after each computer-based training session. The students received a mathematical word problem test, which incorporated worked-out examples presented on the computer. They took the mathematical word problem test for forty minutes. They took a paper and pencil test and a mathematical word problem test at this session from February 2012 to December 2013.

In the fourth session, students also had a transfer test after the final training session. Completion of the transfer test took fifteen minutes.

The students in the control condition who were described in Tajika et al. (2012) had the pretest in April. They solved the same problems

of the pretest two months later. Then they had two kinds of mathematical word problems in February as fifth graders, and in December as sixth graders. They also took the transfer test in December as sixth graders.

3 Results

The results are presented in two parts. First, overall results for the training condition and the control condition over two years are analyzed. The results of the control condition were obtained in the early study of Tajika et al. (2012). We used the data of the control condition by Tajika et al. (2012) to compare with those of the training condition (i.e., the self-explanation condition) of the present study. The students in the control condition received the pretest two months later than the training condition.

The students in the training condition are classified into three groups, based on their scores on the pretest and the mathematical word problem tests carried out in the fourth grade. We show the results of their tests in three groups over two years.

3-1 Overall Results of the Two Conditions over Two Years

Table 1 shows the results of the pretest, the mathematical word problem tests across six test periods, and the transfer test for the students in the training condition. Table 1 also shows the results of the pretest, two mathematical word problem tests, and the transfer test for the students in the control condition. The data for the

control condition were obtained in the early study (see Tajika et al., 2012). Both students of the training condition and the control condition had the same pretest, mathematical word problem tests, and transfer test.

As Table 1 indicates, the scores of the pretest were not significantly different between the two conditions. A comparison of the results from the mathematical word problem tests is shown in Table 1. The students in the training condition solved more word problems correctly than those in the control condition in February of the fifth grade ($t(80)=2.43$, $p<.05$). However, there was no difference of the scores in December of the sixth grade between the two conditions ($t(80)<1$). There was also no interaction between the two conditions and the three word problem test periods.

The transfer test consisted of 18 problems. The score given to a problem was one point, when students generated a correct description. As a result, the maximum score of the transfer test was 18 points. The results of the transfer test are also shown in the right column of Table 1. There was a significant difference between the training condition and the control condition ($t(80)=2.01$, $p<.05$). Students in the training condition performed better than those in the control condition.

3-2 Analyses of Students in the Training Condition over Two Years

According to the typical patterns of performance on the pretest, and the first and second mathematical word problem tests for the

Table 1 Mean Proportion Correct (Ms) and Standard Deviations (SDs) for Two Conditions Over Two Years

	Pretest	Math Word Problem Test						Transfer Test
		Feb 4th	Jul 5th	Dec 5th	Feb 5th	Jul 6th	Dec 6th	Dec 6th
Condition	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Training (n=20)	.86(.18)	.80(.12)	.59(.17)	.80(.18)	.78(.15)	.53(.22)	.53(.23)	.70(.14)
Control (n=62)	.85(.21)				.67(.24)		.56(.21)	.62(.22)

Note. The pretest of the control group was carried out in April after the pretest of the training group had been carried out in February.

fourth graders, we classified the fourth grade students in the training condition into three groups. Each group had six or seven students. Table 2 shows the results of the students for three groups over two years. As shown in Table 2, students in the upper group ($n=7$) kept attaining greater gains over two years. Students in the middle group ($n=7$) and the lower group ($n=6$) were also gradually improving their scores in each grade.

There was a significant difference on the pretest among the three groups, $F(2, 17)=4.18$, $p<.05$. $\eta^2=.41$. Post-hoc HSD tests (p -values $<.05$) revealed that students in the upper group and in the middle group performed better than those in the lower group did.

Next, a 3(Group) \times 6(Test Period) analysis of variance (ANOVA) was conducted on the data of mathematical word problem tests in Table 2. Scores of mathematical word problem tests conducted across six test periods were used for the analysis. There was a significant main effect for group ($F(2, 17)=7.77$, $p<.01$, $\eta^2=.48$). Post-hoc HSD tests revealed that students in the upper group outperformed those in the lower group. There was no difference of scores between the upper group and the middle group. There was also no difference of scores between the middle group and the lower group. There was no significant main effect of test period nor interaction. Choice items for solution steps were recorded as a protocol of their learning history. A detailed protocol analysis showed that there was a high correlation between mean numbers of correct choice items and mean scores on word problem

tests, $r=.78$ ($p<.01$). The mean numbers of correct choice items for worked-out examples which students in the upper group selected were 9.20 ($SD=2.77$). The mean numbers of correct choice items for worked-out examples which students in the middle group selected were 8.20 ($SD=1.30$). The mean numbers of correct choice items for worked-out examples which students in the lower group selected were 6.80 ($SD=2.59$). An ANOVA showed that there was no difference among the three groups ($F(2,15)=1.69$).

Second, we conducted a detailed protocol analysis for each group examining the type of self-explanation in the paper and pencil test that students engaged after the computer-based support activities. Students self-explained each solution step of one worked-out example using paper and pencil during each session of mathematical word problem tests. Students were instructed to self-explain a word problem solution at each step and did not receive feedback on their explanations.

Self-explanations were broken down into inferences versus repetitions and/or others based on the statements of self-explanations of worked-out examples using a paper and pencil test. Self-explanation by inference means that students generate new pieces of knowledge not explicitly stated in each step that are related to the problem solution. For example, a student in the upper group self-explained a solution step for one of the difficult worked-out examples as follows: It takes 10 minutes for the A faucet to fill the tank. So, I think one minute for the A faucet means 1/10 of 10 minutes. The ratio at which the tank is filled up with water is 1/10. We classified such

Table 2 Mean Proportion Correct (Ms) and Standard Deviations (SDs) for Three Groups Over Two Years

	Pretest	Word Problem Test						Transfer Test
	Feb 4th	Feb 4th	Jul 5th	Dec 5th	Feb 5th	Jul 6th	Dec 6th	Dec 6th
Group	M (SD)	M (SD)	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)
Upper ($n=7$)	.82(.20)	.82(.09)	.74(.12)	.95(.10)	.87(.08)	.73(.17)	.62(.23)	.76(.10)
Middle ($n=7$)	.91(.12)	.84(.07)	.47(.16)	.79(.14)	.79(.20)	.47(.19)	.56(.27)	.62(.14)
Lower ($n=6$)	.65(.16)	.73(.18)	.57(.13)	.65(.18)	.67(.08)	.36(.14)	.40(.12)	.71(.16)

explanation as inferential self-explanation.

Most of the students repeated the statements of the worked-out examples and/or gave no responses. The results showed that the number of students in the upper group who self-explained their solution processes using inferences was more than those in the other two groups.

4 Discussion

The result in the present study showed that students in the training condition performed better than those in the control condition on the February scores of the fifth grade. The result also showed that students in the training condition performed superior to those in the control condition on the transfer scores of the sixth grade. However, there were no differences on the other word problem test scores between the two conditions. When students were divided into three groups, the upper group, the middle group, and the lower group, there was a significant main effect for group. The result of the score patterns of the three groups result was the same as that of Tajika et al. (2012). While the students in the upper group used self-explanations as a metacognitive strategy to make incomplete understanding complete, the students in the lower group did not go much beyond the solution steps displayed on the computer screen and on the paper and pencil test.

When students read a solution step to a word problem on a computer screen, they have to activate their schemas to understand the sentence of the solution step. Then to understand the meaning of the sentence they sometimes use inferences. As a result, they self-explain the sentence of the solution step appropriately. They usually construct new knowledge to solve the word problem by self-explanation, making it a metacognitive strategy.

Most fourth-grade elementary school students can self-explain the meaning of each sentence by using Japanese word processors on a computer, but some cannot. We used two to four choice items when students self-explained a sentence of a

solution step. Many of the choice items were selected based on results from other students, and were included in each solution step of worked-out examples. In our study, students selected a correct self-explanation item among choice items, but they did not self-explain the meaning of a sentence. As a result, the performance of students in the training condition on many mathematical word tests was similar to those in the control condition.

Students received a self-explanation test in each test session, called a paper and pencil test of worked-out examples. In our previous study (Tajika et al., 2007), we used paper and pencil tests. Students first self-explained both easy and difficult worked-out examples. Then, they solved ratio word problem tests. Students who self-explained the worked-out examples performed superior to those in the control group.

In this study, students in the training condition self-explained each solution step of a worked-out example using a paper and pencil test after the computer-based training session. However, they self-explained only one easy worked-out example under the limitation of the class. As Tajika et al. (2007, 2012) showed, easily worked-out examples do not promote self-explanation. Students have little chance to infer solution steps for easy worked-out examples. It is possible that this observation helps explain the word problem test results of the training condition.

Self-explanations may usually promote students' word problem solving. However, elementary school children, even fifth and sixth graders, do not spontaneously self-explain solution steps for a worked-out example. In addition, the students had 3-two-week training sessions in a year. The students were asked to solve two or three worked-out examples for twenty minutes on the computer. These seem to be one of the reasons why students in the training condition did not perform better than those in the control condition in every test session. However, the present study has demonstrated that using a computer as a tool by which students can easily self-explain each solution step of a worked-out example is partly effective.

We used the longitudinal approach to collect the mathematical word problem solving data of elementary school children over two years. The students were divided into three groups. We observed students' changes and stabilities in solving word problems. The students in the upper group proved to be quite stable about self-explanation and their word problem solving scores over this year. The students in the lower group improved their word problem solving over time and obtained the scores on the transfer test that were almost as high as those in the upper group. The students in the middle group proved to change their word problem solving ability gradually.

The present study lacked an empirical control condition, in which students have no computer-based metacognitive strategy support. As stated above, the results of the control condition were obtained in the early study of Tajika et al. (2012). It was in April that Tajika et al. carried out the pretest of the control condition. So, the students in the control condition had just become fifth graders and received the pretest two months later than those in the training condition. However, when we compared the present results and those of the control condition of Tajika et al., the present results regarding both the same mathematical word test and the transfer test were significantly better than those in the control condition of Tajika et al., despite the fact that the students were from a younger cohort.

Overall, these findings suggest that many students constructed mental models required for solving mathematical word problems by using self-explanation, which in turn led to enhanced performance for these students.

5 Acknowledgements

The study was supported by a Grant-in-Aid for Scientific Research (c), No.23530881 and a Grant-in-Aid for Scientific Research (c), No.26530881, from the Japan Society for Promotion of Science.

6 References

- Aleven, V., & Koedinger, K.R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26, 147-179.
- Azevedo, R., & Allevén, V. (2013). *International handbook of metacognition and learning technologies*. New York: Springer.
- Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13, 221-252.
- Chi, M.T.H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R.Graser (Ed.), *Advances in instructional psychology* (Vol. 5, pp. 161-238). Mahwah, NJ: Erlbaum.
- Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Chi, M.T.H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49, 219-243.
- Chiu, J.L., & Chi, M.T.H. (2014). Supporting self-explanation in the classroom. In V.A.Benassi, C.E.Overson, & C.M.Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the classroom* (pp. 91-103). Retrieved from the Society for the Teaching of Psychology web site: <http://teachpsych.org/ebooks/esle2014/index.php>.
- Flavell, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.
- Mayer, R.E., Tajika, H., & Stanley, C. (1991). Mathematical problem solving in Japan and the United States: A controlled comparison. *Journal*

- of *Educational Psychology*, 83, 69-72.
- Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful?: The case of analogical problem-solving. *British Journal of Educational Psychology*, 68, 15-24.
- Neuman, Y., & Schwarz, B. (2000). Substituting one mystery for another: The role of self-explanations in solving algebra word-problems. *Learning and Instruction*, 10, 203-220.
- Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Schoenfeld, A.H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.
- Schoenfeld, A.H. (1987). What's all the fuss about metacognition? In A.H.Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189-215). Hillsdale, NJ: Erlbaum.
- Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp.334-370). New York: MacMillan.
- Tajika, H., Nakatsu, N., Neumann, E., Nozaki, H., Kato, H., Fujitani, T., & Hotta, C. (2012). Mathematical word problem solving in children engaged in computer-based metacognitive support: A longitudinal study. *Educational Technology Research*, 35, 11-19.
- Tajika, H., Nakatsu, N., Nozaki, H., Neumann, E., & Maruno, S. (2007). Effects of self-explanation as a metacognitive strategy for solving mathematical word problems. *Japanese Psychological Research*, 49, 222-233.

7 Note

The study was also presented in a poster session at the 7th Biennial SIG Meeting of Metacognition from August 23-26, 2016, Nijmegen, the Netherlands.